Glossary: Working with Data in Python

Welcome! This alphabetized glossary contains many of the terms you'll find within this course. This comprehensive glossary also includes additional industry-recognized terms not used in course videos. These terms are important for you to recognize when working in the industry, participating in user groups, and participating in other certificate programs.

Term	Definition
.csv file	A .csv (Comma-Separated Values) file is a plain text file format for storing tabular data, where each line represents a row and uses commas to separate values in different columns.
.txt file	A .txt (Text) file is a common file format that contains plain text without specific formatting, making it suitable for storing and editing textual data.
Append	To "append" means to add or attach something to the end of an existing object, typically used in the context of adding data to a file or elements to a data structure like a list in Python.
Attribute	An "attribute" in Python refers to a property or characteristic associated with an object, which can be accessed using dot notation.
Broadcasting in NumPy	Broadcasting in NumPy allows arrays with different shapes to be combined in element-wise operations by automatically extending smaller arrays to match the shape of larger ones, making operations more flexible.
Component	In NumPy, a "component" typically refers to a specific element or value within a multi-dimensional array, which can be accessed using indexing.
Computation	Computation in NumPy involves performing numerical operations on arrays and matrices, making it a powerful library for mathematical and scientific computing in Python.
Data analysis	Data analysis is the process of inspecting, cleaning, transforming, and interpreting data to discover useful information, draw conclusions, and support decision-making.
DataFrames	A DataFrames in Pandas is a two-dimensional, tabular data structure for storing and analyzing data, consisting of rows and columns.
Dependencies	Dependencies in Pandas are external libraries or modules, such as NumPy, that Pandas rely on for fundamental data manipulation and analysis functionality.
File attribute	File attributes generally refer to properties or metadata associated with files, which are managed at the operating system level.
File object	A "file object" in Python represents an open file, allowing reading from or writing to the file.
Grid	In Python, a "grid" typically refers to a two-dimensional structure composed of rows and columns, often used to represent data in a tabular format or for organizing objects in a coordinate system.
Hadamard Product	The Hadamard product is a mathematical operation that involves element-wise multiplication of two matrices or arrays of the same shape, producing a new matrix with each element being the product of the corresponding elements in the input matrices.
Importing pandas	To import Pandas in Python, you use the statement: import pandas as pd, which allows you to access Pandas functions and data structures using the abbreviation "pd."
Index	In Python, an "index" typically refers to a position or identifier used to access elements within a sequence or data structure, such as a list or string.
Libraries	Libraries in Python are collections of pre-written code modules that provide reusable functions and classes to simplify and enhance software development.
Linespace	In Python, "linespace" refers to a NumPy function that generates an array of evenly spaced values within a specified range.
NumPy	NumPy in Python is a fundamental library for numerical computing that provides support for large, multi-dimensional arrays and matrices, as well as a variety of high-level mathematical functions to operate on these arrays.
One dimensional NumPy	A one-dimensional NumPy array is a linear data structure that stores elements in a single sequence, often used for numerical computations and data manipulation.
Open function	In Python, the "open" function is used to access and manipulate files, allowing you to read from or write to a specified file.
Pandas	Pandas is a popular Python library for data manipulation and analysis, offering data structures and tools for working with structured data like tables and time series.
Pandas library	Pandas library in Python refer to the various modules and functions within the Pandas library, which provides powerful data structures and data analysis tools for working with structured data.
Plotting Mathematical Functions	Plotting mathematical functions in Python involves using libraries like Matplotlib to create graphical representations of mathematical equations, aiding visualization, and analysis.
Shape	In NumPy, "shape" refers to an array's dimensions (number of rows and columns), describing its size and structure.
Slicing	Slicing in NumPy entails extracting specific portions of an array by specifying a range of indices, enabling you to work with subsets of the data.
Two dimensional NumPy	A two-dimensional NumPy array is a structured data representation with rows and columns, resembling a matrix or table, ideal for various data manipulation and analysis tasks.
Universal Functions	Universal functions (ufuncs) in NumPy are functions that operate element-wise on arrays, providing efficient and vectorized operations for a wide range of mathematical and logical operations.
Vector addition	Vector addition in Python involves adding corresponding elements of two or more vectors, producing a new vector with the sum of their components.
Visualizations	Visualizations in Python refer to the creation of graphical representations, such as charts, plots, and graphs, to illustrate and communicate data and trends visually.